Annals of Oncology 16 (Supplement 8) viii20–viii29, 2005
doi:10.1093/annonc/mdi963

Clinical trials in ovarian carcinoma: study methodology

J. B. Vermorken[1]*, M. K. B. Parmar[2], M. F. Brady[3], E. A. Eisenhauer[4], T. Hogberg[5], R. F. Ozols[3], J. Rochon[6], G. J. S. Rustin[2], S. Sagae[7] & R. H. M. Verheijen[1]

Gynecologic Cancer Intergroup (GCIG) and its member organizations: [1]EORTC (Europe); [2]MRC/NCRI (UK); [3]GOG (USA); [4]NCIC-CTG (Canada); [5]NSGO (Scandinavia); [6]AGO-OVAR (Germany); [7]JGOG (Japan)

Introduction

One of the major issues in the 3rd Ovarian Cancer Consensus Conference (OCCC) was to achieve consensus among study groups worldwide on appropriate requirements for entry criteria and end points for clinical trials in ovarian cancer. A 'clinical trial' was defined as a carefully designed, prospective medical study that attempts to answer a precisely defined set of questions with respect to the effects of a particular treatment or treatments [1]. The focus was primarily on phase II and III trials: the goal of the latter is to determine either the effectiveness of a treatment relative to the best current standard of care or whether a new treatment is as effective as a standard, but associated with less toxicity, cost or better quality of life. The design, execution and analysis of phase III trials not only should be based on sound scientific and ethical criteria, but it was agreed by all attendees that such trials must have sufficient statistical power to undertake an analysis of survival [2]. Historically, inadequately powered trials have undermined our ability to draw reliable conclusions on the values of different treatment approaches. As a result, several important questions remain the subject of continuing debate, despite randomized studies, including the exact role of chemotherapy in patients with high-risk early ovarian cancer after comprehensive surgical staging, the optimal number of treatment cycles in the treatment of advanced disease, the role of maintenance therapy and/or consolidation therapy, and the usefulness of dose-dense therapies and high-dose chemotherapy with autologous stem cell support. The future will be even more demanding with the evaluation of new drugs aimed at an ever increasing number of molecular targets [3]. For these reasons a worldwide consensus on standards for trials, particularly randomized studies, seems to be very timely.

Of the 12 questions that were addressed during the OCCC, three concerned study methodology and are the subject of this paper. These questions were as follows (Table 1).

Table 1. Consensus questions addressing the topic 'study methodology'

1. Which patient/disease characteristics should be considered as entry criteria or at least as strata for subgroup analysis in trials?
2. Which kind of phase III randomized study design can be recommended to the study groups to make future trials quicker, cheaper and more reliable?
3. Which are the recommended primary end points for future phase II and randomized phase III clinical trials in ovarian cancer?

Question 1. Inclusion criteria for ovarian cancer clinical trials: with focus at strict versus broad eligibility (ICON-like) criteria

In defining inclusion criteria for trials, one must consider whether certain baseline disease or patient factors lead to sufficiently different outcomes such that differing treatments or trials are appropriate. For these reasons, ovarian cancer studies have been conducted in three broad separate settings: front-line therapy in early disease, front-line therapy in advanced disease (as defined below) and therapy in recurrent disease. However, even within these categories, often clinical and pathological factors have been shown to have prognostic impact. For *advanced* ovarian cancer (FIGO stages IIB–IV) the 2nd OCCC (1998, Bergen aan Zee, The Netherlands) recommended that for adequate analysis the following details of known prognostic importance should be recorded on patients who were entered on front-line studies: age, performance status, histology, tumor grade (degree of differentiation), stage and residual disease (microscopic or none versus macroscopic) [4]. Entry criteria usually specify the limits of eligibility around these parameters.

In contrast, the two most important prognostic factors in patients with *early* ovarian cancer (FIGO stages I–IIA) are the degree of differentiation (grade of the disease) and the completeness of staging [5, 6]. However, stage, extracapsular growth, spontaneous rupture, the presence of ascites, DNA ploidy or DNA index (a quantitative pathology measure) and elevated CA 125 have also been identified as independent prognostic factors in some multivariate analyses and thus some of these are often specified as part of entry criteria [7, 8].

Most studies in patients with *recurrent* disease have written entry criteria based on those factors predictive of response to treatment rather than on survival. Time since last chemotherapy

**Correspondence to:* Dr J. B. Vermorken, Department of Oncology, University Hospital Antwerp, 2650 Edegem, Belgium.
Tel: +32 3-821-39-54. Fax: +32 3-825-05-64.
E-mail: Jan.b.Vermorken@uza.be

has been the most commonly used measure to predict the likelihood of response to second-line therapy and many trials segment or restrict the population according to pre-specified time periods [7]. However, a large meta-analysis of several second-line chemotherapy studies (using data from >700 patients) indicated that other factors could also play a role in determining response, such as disease bulk, number of disease sites involved and histology [9]. Using the same dataset, significant factors at the start of second- or third-line therapy for longer subsequent survival were longer time since diagnosis, longer time since last chemotherapy, better performance status, low disease bulk, histology (non-mucinous), fewer disease sites involved and a normal hemoglobin level [10].

These data suggest that different prognostic groups can be identified even within the three traditional categories and that perhaps cohorts of patients defined by these criteria should be treated differently. This reasoning could be used as justification to use more restrictive eligibility criteria and to perform trials in multiple smaller subsets of patients. It should be understood, though, that even when trial entry is restricted, heterogeneity in the types of patients actually entered will take place and that, if too narrow a population is stipulated, the trial results may not be generalizable to the entire population. Furthermore, even if the plan is to be reasonably restrictive in patient entry for the purposes of being able to make comparisons across trials, there are problems in doing so. The following examples highlight these points.

(a) In *early* ovarian cancer several recent studies have focused on the role of chemotherapy in the so-called high-risk disease setting (ICON1, ACTION, GOG#157 [11–14]). Both ICON1 and ACTION compared platinum-containing adjuvant chemotherapy versus observation following surgery. In both studies the primary end point was survival. However, the entry criteria in both trials were different. In ICON1 these were quite liberal, i.e. any patient in whom the clinician was uncertain whether that patient should receive chemotherapy could enter the trial and surgery primarily consisted of total hysterectomy and bilateral salpingo-oophorectomy [13]. In the ACTION trial, however, the entry criteria were more restrictive, i.e. only patients with FIGO stages IA, IB (grades II and III), stages IC and IIA (all grades) and all clear cell carcinomas could enter the trial. Furthermore, ACTION had more strict guidelines for surgery consisting of total abdominal hysterectomy and bilateral salpingo-oophorectomy, followed by surgical staging, as indicated in the EORTC surgical guidelines [14]. The combined analysis on 925 patients showed an 8% improvement in survival for immediate chemotherapy versus observation, and despite differences in eligibility criteria (liberal or restricted), staging requirements and chemotherapy (more single-agent carboplatin in ICON1, more cisplatin-based combinations in ACTION), the individual results of the two trials were very similar, with the magnitude of the effect of chemotherapy being of very similar size. Subgroup analysis failed to demonstrate a different effect of chemotherapy in any of the subgroups that could be analyzed, i.e. age, tumor stage, histological cell type and grade of differentiation [11]. Unfortunately, the relationship between staging performance and the effect of chemotherapy could not be adequately analyzed. Only one-sixth of the population was optimally staged and, despite the more strict staging requirements in ACTION, only one-third of the patients received proper staging. A separate analysis of the ACTION trial suggested that the benefit of adjuvant chemotherapy was limited to patients with non-optimal staging [14]. However, there was not enough statistical power to deny a positive effect in patients who had been optimally staged. So, the interpretation remains unclear leading to a variety of attitudes in different countries. In GOG#157, a trial that studied the impact of longer versus shorter duration of adjuvant chemotherapy, staging was required as per GOG published guidelines [12]. However, of the 457 patients, only 70% met all eligibility criteria and 23% (107/457) were incompletely staged. These data suggest that in daily practice in non-specialized centers the percentage of patients with optimal staging will be substantially lower. So, optimal staging in early ovarian cancer remains problematic and the relevance of it to patients treated outside of clinical trials might be even more questionable.

(b) In the *advanced* disease setting similar difficulties have been encountered. Examples of this are the remarkable differences in outcome between protocol GOG#111 and protocol GOG#132, trials which were performed in sequence by the GOG in patients with the same eligibility criteria (suboptimal stages III and IV), and using the same treatment (paclitaxel 135 mg/m^2, 24 h, plus cisplatin 75 mg/m^2 every 3 weeks for six cycles). Progression-free and overall survival were 18 and 38 months, respectively, in GOG#111, and 14 and 27 months, respectively, in GOG#132 [15, 16]. So, even though the eligibility criteria were the same, there must have been a selection bias with worst prognosis patients in GOG#132. This means that other criteria besides stage and volume are important and need to be identified.

The assessment of the amount of residual disease is a particularly difficult item and open to much variation in subjective interpretation. There are now at least three large randomized trials showing that progression-free or overall survival are improved when cisplatin-based intraperitoneal (i.p.) chemotherapy is applied compared with intravenous administration of platinum-based chemotherapy [17–19]. The fact that in the first positive trial (the purest of all) no statistical significant advantage for intraperitoneal chemotherapy was found in the subset of patients with <0.5 cm disease is still puzzling and has led to a negative interpretation by some and has reopened the debate as to which patient population will ultimately benefit from i.p. therapy. How standard and objective are the methods of measuring the size of the largest residual peritoneal mass in the operating room? Maybe the distinction between suboptimally and optimally debulked disease should indeed be made on the basis of macroscopic versus no macroscopic disease left after surgery and i.p. therapies should be further studied in the latter category.

Clearly it seems very difficult, if not impossible, to draw reliable conclusions when comparing across different clinical trials, however similar they appear to be. This makes it even more important that comparisons within trials are as reliable as possible, which in turn emphasizes the overriding need for large-scale studies whenever possible.

(c) A further negative effect of having very strict eligibility criteria is that it may lead to slow accrual. A clear example of this is EORTC protocol 55875, a randomized phase III study in ovarian cancer patients with a pathologically complete remission after platinum-based intravenous chemotherapy [20]. The study evaluated the role of i.p. cisplatin versus no further treatment. It took 9 years to accrue 153 patients and the study was closed prematurely, and suffered from the inclusion of women who were either ineligible (10%) or had major protocol violations (11%), part of which were most likely related to limited experience with the technical aspects of i.p. therapy. Moreover, there was an awareness of a progressive change in 'standard' first-line intravenous chemotherapy, with paclitaxel–cisplatin progressively replacing cyclophosphamide–cisplatin. Nevertheless, the trial showed the same trend as the other i.p. trials, i.e. a superior outcome.

These examples gave sufficient room for discussion on how strict or how flexible one should be with respect to eligibility criteria. The less restrictive (ICON-like) eligibility criteria seem more in line with what is applicable to the general population; with this approach more patients can be entered and accrual is facilitated. The volume of the residuum in stage III disease may be a biased criterion when it concerns measurement of residual disease; however, as mentioned earlier, further studies in patients with no residual macroscopic disease may overcome this bias.

Therefore, working group B concluded that there are no hard and fast rules as to which types of patients should be entered into phase III clinical trials. However, some considerations need to be taken into account.

We noted that the first randomized trial of a new therapy is often carried out, partly as matter of expediency, in patients with stage IV or recurrent disease, where there is a high event rate and thus results come more quickly. If results are positive, this has sometimes led to further trials in earlier stages of the disease. We need to be aware that such a model might be appropriate, but might also be misleading. For example, 5-fluorouracil is not very active in advanced colon cancer, but has now become a mainstay of adjuvant treatment of this disease.

In any framework that considers inclusion and exclusion criteria for trials in ovarian cancer, it is important to consider not only who should be *included* in any given trial, but also whether any particular subgroup should be *excluded*. To address this it is useful to consider the following three questions:

a) *Is the prognosis of the subgroup of patients sufficiently different to the group as a whole to conclude without further information that it is inappropriate to include this group of patients?*

As an example, it is very unlikely that we would include stage IA grade I patients in the same trial as stage IV patients. This is mainly because the prognosis of these two groups of patients is so different that it is very unlikely that after surgery we would want to follow similar treatment strategies for them. However, the same argument may not hold for stage III and IV patients, whose prognosis although different, may not be different enough to *a priori* entertain different treatment strategies.

b) *Is there 'good' biological, medical or statistical evidence that the treatment is going to be considerably more or less effective (or even ineffective) in a particular subgroup of patients?*

As an example, we know that many therapies are likely to be more effective in patients with recurrent disease who have platinum-sensitive disease than in patients with platinum-refractory disease. Thus for most new therapies we would not include both groups of patients in the same trial. An exception would be, if there are strong preclinical data that an agent may be active only when a specific biologic marker is present, to include ovarian cancer patients with tumors that overexpress the marker and to exclude those whose tumors do not. An example of this is testing trastuzumab only in ovarian cancer patients with measurable persistent or recurrent epithelial ovarian cancer with 2+ or 3+ HER2 overexpression [21].

c) *Is it likely that any result from the trial will be generally extrapolated to include a particular group of patients?*

As an example: we know that if we perform a trial in stage IV disease, that in many cases any result is likely to be extrapolated to patients with stage III disease. In this case it would have been better to also include the stage III patients into the trial to assess whether there is evidence of a different size of effect in stage III and IV patients. Another common example is that of elderly patients who should most probably not be excluded from trials evaluating standard chemotherapy regimens, because it is very likely that results of such trials will be generalized to this patient population also.

To summarize, if the answers to questions (a) and (b) are 'no', then strong consideration should be given to *including* the subgroup of patients in the trial. Whatever the answers to (a) and (b), if the answer to (c) is 'yes' then, again, consideration should be given to *including* this group of patients into the trial. Thus, the answer to the first question on study methodology is as follows, which after discussion with the whole consensus panel obtained a unanimous acceptance (Table 2).

Table 2. Consensus statements in response to question 1

Which patient/disease characteristics should be considered as entry criteria or at least as strata for subgroup analysis in trials?
The following patient/disease characteristics should be formally considered for patients entry or as stratification factors:
Primary site, stage, prior treatment history, histological type, grade, residual disease, measurable or non-measurable disease, serum CA 125, performance status, age and co-morbidity and other validated prognostic factors. For post-recurrence/progression trials: disease-related symptoms and treatment-free interval.
Before exclusion of any particular patient group the following questions should be considered:
Is the prognosis of these patients sufficiently different to the group as a whole to conclude without further information that it is inappropriate to include this group of patients?
Is there good biological, medical or statistical evidence that the treatment is predicted to be considerably more or less effective (or even ineffective) in this group of patients?
Is the result from the trial likely to be applied to this group of patients?

Question 2. Trial design in ovarian cancer clinical trials: with focus at multi-arm versus single-question trials

With the increasing pace of drug development and the pressure to get answers more quickly, it is reasonable to consider the most optimal design(s) for large randomized trials to arrive at answers rapidly and efficiently. The sample size needed for such trials is substantial if modest, but real, overall or progression-free survival differences are to be detected (see later). Trials performed by GCIG groups have been able to accomplish this using the traditional two-arm trials: the median sample size in the five completed GCIG first-line ovarian cancer trials was 1300 patients [see Gynecologic Cancer Intergroup (GCIG): History and current status, this issue] and accrual time for these trials ranged from 2 to 3 years.

A complicating (but fortunate) factor in deliberating efficient trial design is that, at the present time, it is not unusual for *several* promising new agents or treatment regimens to be ready simultaneously for testing in a randomized phase III setting. It may be impractical and inefficient to test several new therapies in individual trials against a control arm by conducting multiple trials using a conventional parallel two-group design. For example, too few patients may be available given the required sample size of each of these trials, or the resources needed (for example, the costs) may be too great [22]. On the other hand, performing such trials sequentially would take too much time. Therefore, novel multi-arm designs in which a control regimen is compared with several new (experimental) therapies are worth considering. The issue of using multi-arm trials or single question studies was extensively debated in working group B and led to the recommendations that can be found at the end of this section in Table 4. These recommendations were in turn accepted unanimously by other Workshop representatives.

There are pros and cons to conducting a single multi-arm trial versus several two-arm studies. Multi-arm trials can be considerably more complex to design, conduct and analyze than two-arm, single-question trials. The additional complexities can be classified as arising from ethical, administrative or scientific/statistical considerations.

Ethical challenges

Obtaining the patient's informed consent for multi-arm trials is more challenging than the simpler two-arm trial when treatment arms include a broad range of agents. Since the patient's consent must be based on making an informed decision, additional care is required to ensure that prior to enrolling onto the study the patient understands the detailed information concerning the risks associated with each of the study regimens of which only one will ultimately be administered.

Administrative challenges

Multi-arm phase III clinical trials in gynecologic malignancies are likely to require collaboration among multiple cooperative groups. For example, the five-arm advanced ovarian cancer trial, GOG-182/ICON5, involved cooperative groups from Australia, New Zealand, Italy, UK and the USA. In this case, each of these groups had prior experience and established procedures for conducting phase III trials; however, collaboration requires standardization of these procedures. Each group makes concessions in order to develop uniform procedures for study development, conduct and monitoring. Standardizing the data monitoring process requires identifying those clinical observations that are necessary to meet the study objectives and developing a common set of data forms and data definitions that can be unambiguously implemented across all treatment centers and data centers involved in the study.

Multi-national studies introduce additional unique challenges. It is not uncommon for investigational agents to be available in some countries but not in others. Moreover, the regulatory procedures enforced within each country are not universal and individual investigators are often unable to make concessions in order to promote study-wide standards. Indeed, laws and regulations in each country are not static and therefore, procedures that are apparently sufficient at the initiation of the study may require modifications before the trial is completed. Furthermore, if all or several agents to be studied are investigational, competing pharmaceutical firms may not agree to have their agents studied in the same trial for business reasons, regardless of the scientific merit of the proposal. This latter situation may call for considerable negotiating skills.

It is reasonable to expect that future multi-arm trials will require even greater organizational efforts if they include investigational agents and require direct involvement of the industry. Trial sponsors from industry will typically impose additional study objectives and constraints on the conduct and administration of the study beyond those deemed appropriate for scientific reasons.

The eligibility criteria for multi-arm trials may also be more restrictive than for two-arm trials. Each additional treatment arm may either increase the requirement for restricting eligibility or reduce the patients' interest in participating in the study. For example, trials with a regimen containing an anthracycline may make it necessary to limit eligibility to patients who have not recently experienced congestive heart failure. Trials with a taxane regimen may eliminate patients experiencing peripheral neuropathy. These eligibility criteria that are considered justifiable for safety's sake have the unfortunate cumulative impact on reducing the number of patients who can participate in the trial. Moreover, some of the otherwise eligible patients may not be willing to accept randomization to all of the study treatments. For example, in a recent multi-arm trial evaluating tamoxifen and radiotherapy for the treatment of ductal carcinoma *in situ* of the beast, 46% of the eligible patients were willing to have either radiation or tamoxifen treatment randomly assigned, but not both [23]. To some extent these eligibility restrictions and patient preferences can be mitigated in multi-arm trials by using more complex randomization and analytic procedures [24].

Scientific and statistical challenges

The scientific challenges of multi-arm trials stem from the increased number of hypotheses that can be tested. In a two-arm trial there is only one treatment comparison; however, in a clinical trial involving k different treatments, there are potentially $k(k - 1)/2$ pair-wise treatment comparisons. That is, in a trial with five treatment arms there are potentially 10 distinct pair-wise treatment comparisons. Suppose that all five of the treatment regimens are truly equivalent and at the end of the study each pair-wise treatment comparison is tested at the traditional 0.05 significance level. In this case the probability of *incorrectly* declaring at least one treatment to be superior to another is 23% (Table 3). Such a high probability for this type of error is usually considered too great for a phase III trial. Typically, phase III trials control this error (called type I error) so that it does not exceed 5%.

There are several approaches that can be considered for limiting type I errors in multi-arm trials. The first approach is to require a greater level of evidence before declaring two treatment regimens different. For example, rather than requiring a P value <0.05 in order for a difference to be considered statistically significant, a trial could require P values to be $<0.05/m$, where m is the number of planned treatment comparisons. This adjustment is commonly called the Bonferroni procedure [25]. Therefore, in a five-arm trial in which all 10 pair-wise treatment comparisons are planned, requiring the P value to be $<0.05/10 = 0.005$ will limit the study-wide probability of type I error to no more than 5%. While the Bonferroni adjustment is easily applied, this procedure also reduces the chance of detecting differences between treatments when they truly exist (statistical power). There are other adjustment procedures that can be used to control type I errors in multi-arm trials that are slightly more complicated but preferable because they are more likely to detect differences when they truly exist [26]. All these adjustment procedures improve the specificity of the trial (reduce the probability of a type I error). However, without a corresponding increase in the size of the trial, these adjustment procedures also reduce the sensitivity of the trial for detecting differences between treatments when they truly exist. Therefore, multi-arm trials typically enrol more patients onto each treatment arm than a similarly designed two-arm trial in order to improve sensitivity while controlling overall specificity.

When there are fewer treatment comparisons made, there are fewer opportunities to make an error. This suggests a second approach for limiting type I errors within a multi-arm trial by limiting the number of planned treatment comparisons. This approach may not be as undesirable as it first appears. Consider a multi-arm trial in which one of the study treatments is the standard intervention. Also, suppose that there is an *a priori* preference for the standard treatment. In other words, the standard treatment will continue to be recommended unless the trial provides overwhelming evidence indicating that at least one of the experimental regimens is significantly better than the standard intervention. In this type k-arm trial, there are only $(k - 1)$ comparisons between the standard treatment group and each of the experimental treatment groups that are of immediate interest. No comparisons between the pairs of experimental treatments are planned unless one experimental regimen is deemed superior to the control arm. Therefore, the Bonferroni-adjusted critical P value for this five-arm trial is $0.05/4 = 0.0125$, rather than 0.005 as in the previous five-arm trial described earlier. If all of the treatments approaches in this five-arm study are truly equivalent, this approach limits the probability of incorrectly accepting an experimental treatment as the new standard of care to no more than 5%. In order to maintain sensitivity this approach also requires increasing the number of patients enrolled. While the number of patients to be enrolled onto each treatment arm is still larger than that required for a two-arm trial, the increase is not as large as the multi-arm trial that does not restrict the number of treatment comparisons.

It is reasonable to wonder why an investigator who plans to compare several new experimental regimens to a standard treatment in a single multi-arm trial should use statistical considerations different from the investigator who decides to study the same regimens using several sequential two-arm trials. The difference between these two approaches arises from the dependence among the treatment comparisons when a single multi-arm trial is performed. Consider a single multi-arm trial in which the control group includes slightly more patients with a good prognosis than expected. In this trial all of the experimental regimens will tend to appear less beneficial than they truly are. Likewise, if the control regimen happens to include more patients with a poor prognosis than expected, then all of the experimental regimens will appear more active than they truly are. In other words, if one experimental arm in a multi-arm trial is deemed significantly better than the control arm, then it is more likely that another experimental regimen will also be deemed significantly better than the control [27]. There is a dependence among the estimated experimental treatment effects sizes introduced into the design and analysis when all of them are being compared with a single control arm. This dependence does not occur when each experimental regimen is compared to a different control arm, as when there are multiple sequential two-arm trials.

In summary, multi-arm clinical trials have ethical, administrative and scientific considerations that may not be present in two-arm trials. An ethical challenge can arise from the necessary information that a patient needs to understand regarding several experimental regimens in order to make an informed consent prior to enrolling onto the trial. Administrative challenges may arise from the need for greater resources required for conducting multi-arm trials. The greater scientific challenge in multi-arm trials is due to the proliferation of study objectives. There is no

Table 3. Number of possible comparisons and the probability of erroneously declaring one or more treatments different (type I error) in a multi-arm trial when $\alpha = 0.05$ for each test and there is no adjustment for multiple comparisons

Number of treatment groups	Number of possible pair-wise comparisons	Probability of at least one type I error in the entire study
2	1	0.050
3	3	0.113
4	6	0.178
5	10	0.234

Table 4. Consensus statements in response to question 2

Which kind of phase III randomized study design can be recommended to the study groups to make future trials quicker, cheaper and more reliable?
There is a continuing need to conduct large scale randomized trials requiring international collaboration through the GCIG.
The primary determinants for whether to use multi-arm or two-arm designs are study objectives, prioritization of the clinical questions and the availability of resources.
When questions to be answered are of similar priority, multi-arm trials may be preferable.

longer a single alternative hypothesis in multi-arm studies. Adjustments should be made for multiple correlated estimates.

Question 3. Relevant end points for clinical trials in ovarian cancer

The first and most important step in planning a clinical trial is to indicate clearly the primary and secondary objectives [2]. What questions is the trial being designed to answer? Once the objectives are known, this identifies the primary and secondary end points of the study. Trial end points can be classified as either 'true' or 'surrogate'. True end points have direct clinical relevance to the patient, such as symptoms improvement, survival duration, or cure rates. Surrogate end points assess events that are in the etiologic pathway to a true outcome [28]. The primary reason for using a surrogate end point instead of a true end point is either to reduce the duration (because this end point occurs earlier than the actual end point) and cost of a clinical trial or if it is believed that salvage therapies may obscure the effect of the study treatment on a true end point. As an example, progression-free survival has often been considered a surrogate end point for overall survival in trials including patients with advanced ovarian cancer. It is noteworthy that the justification for using a particular surrogate end point is frequently based on data suggesting a statistical correlation with a true end point. However, a correlation between a surrogate and true end point is a necessary, but not sufficient condition to justify a particular surrogate end point. The ideal surrogate end point for randomized trials is an intermediate event in the only causal pathway to the true end point, and the effect of an intervention (i.e. a treatment) on the true end point should be through its influence on the surrogate end point [2]. Reasons for failure of a surrogate end point could be explained in several ways: i.e. either (i) of several causal pathways of disease, the intervention only affects the pathway mediated through the surrogate, or (ii) the surrogate is not in the pathway of the intervention's effect, or is insensitive to its effect.

In ovarian cancer trials the traditional patient specific outcomes of interest often include: overall survival (or cure rate) and progression-free survival, response and toxicity, and symptom control/quality of life. Of these there is general agreement on overall survival (or cure rate) and symptom improvement/quality of life as primary meaningful end points [although quality of life is not (yet) used as such] and toxicity is considered a necessary measure (and primary end point for phase I studies). However, there is debate about the importance of response and progression-free survival as being meaningful end points due to the uncertainty as to whether the patient has any benefit from a longer time to tumor progression or from tumor regression itself. Regardless of the 'meaning' of these end points in and of themselves, if either or both were shown to be true surrogates of survival or quality of life, their use as primary trial end points is easily justified.

Phase II end points

In phase II trials of new agents (or combination) in ovarian cancer, where the primary objective is to determine early evidence of biologic effect of the new drug(s), historically objective response has been defined as the primary end point. It has the advantage of being non-invasive, subject to internationally recognized standardized criteria [29] and readily determined after a series of treatment cycles. Moreover, it is not influenced by salvage therapy. Its disadvantage, though, is the fact that by definition patients must have measurable disease at baseline to be evaluated. While this is usually the case in recurrent disease, it may not be so in the front-line setting. Furthermore, interobserver variability in declaring response, even according to objective measures, has been documented [30]. Because ovarian cancer is often associated with elevation of the well-studied serum antigen, CA 125, and since the levels of the antigen correlated with disease burden, changes in the level of CA 125 seem a plausible substitute for objective tumor regression. Following on work originally conducted by Rustin where a set of CA 125 response criteria were suggested [31], the GCIG Response/Progression Working Group has defined modified Rustin (CA 125) criteria to be used prospectively as an addition to Response Evaluation Criteria in Solid Tumours (RECIST) as a method of defining response in relapsed ovarian cancer patients [29, 32]. The validity of the 50% response definition according to Rustin (later endorsed as the GCIG response criteria) as a substitute for objective response as assessed by RECIST was confirmed by the GINECO group in France in the setting of recurrent disease [33]. Prospective validation of these modified Rustin criteria (GCIG CA 125 definition) in recurrent disease is awaited, and several groups are using these in ongoing trials. For front-line trials, CA 125 response criteria also await validation and therefore cannot be used as such in that setting yet.

Phase III trials

In randomized trials there is no systematic evidence that objective response is a surrogate for overall survival. Furthermore, there are limited data on its surrogate value in assessing quality of life. Nevertheless, it is of interest that quality of life studies in relapsed ovarian cancer patients have indicated that quality of life scores improve in patients who respond to chemotherapy, confirming the palliative nature of chemotherapy [34]. There is obviously an inverse relationship between experienced toxicity and quality of life. This has been observed in randomized phase III trials in the front-line setting, e.g. in trials in which cisplatin was replaced by carboplatin in the combination with paclitaxel

[35]. Such differences have so far not been observed in randomized trials in the recurrent disease setting, but the information about it is scarce. Interestingly, although most consider quality of life an important primary end point for trials in incurable disease settings, it is seldom, if ever, a primary end point in randomized trials in recurrent ovarian cancer.

The GCIG members have accepted the definition of CA 125 progression, in contrast to CA 125 response, as an addition to objective disease progression in front-line randomized trials [36]. A patient may be declared to have progressive disease on the basis of either the objective RECIST criteria or the CA 125 progression criteria. The date of the progression will be the date of progression of the earlier of the two events if both are documented. Since it was recognized that the timing of investigations during first-line therapy and subsequent follow-up may also influence the assessment of progression-free survival in clinical trials, it was proposed that serum CA 125 levels would be obtained on day 1 of each chemotherapy cycle, 4 weeks after the last course, thereafter every 3–4 months for the first 36 months, every 6 months from months 37–60, and every year from 5 years after the primary diagnosis [36]. Although it is recommended that the date of progression is recorded according to both CA 125 and RECIST criteria, it is important to continue validation of CA125 progression by determining whether the trial outcome would be the same whether CA 125 was used or not.

End points: front-line phase III studies

The main issue in discussing randomized phase III front-line studies is whether progression-free survival (for advanced ovarian cancer) or relapse-free survival (for early ovarian cancer) can ever be considered meaningful primary end points. If one agrees that improvement in overall survival is the finding for which we would change our standard of care, then progression-free and relapse-free survival could be considered as alternative primary end points if the available data is strong enough to consider them valid surrogates for survival. What are these data?

For *early ovarian cancer*, there is only one adequately powered trial in the adjuvant setting, the combined ICON1/ACTION analysis [11]. Results from this trial showed that relapse-free survival differences were mirrored in the overall survival analysis. Data from other therapeutic areas such as breast cancer seem to support the use of relapse-free survival as a valid surrogate for survival in the adjuvant setting. Clinicians have changed practice and drugs do get approved for significant improvements in relapse-free survival without waiting for overall survival data [37]. Thus the use of relapse-free survival as a primary end point in randomized trials of adjuvant therapy in early ovarian cancer is justified not only by extrapolation from other solid tumor settings but also by data from the largest randomized trial in early ovarian cancer itself.

The data are also strong for recommendation of progression-free survival as a primary end point in front-line trials in *advanced ovarian cancer* on several counts. Recent adequately powered trials where progression-free and overall survival are known have shown concordant observations between progression-free survival differences and overall survival differences [15, 16, 38–41]. Buyse et al. [42] showed in a meta-analysis of advanced ovarian cancer trials (data from the Ovarian Cancer Meta-analysis Project [43]) that by applying a new method for validation of surrogate end points the treatment effects on the true end point (logarithm of survival) and the treatment effects on the surrogate end point (logarithm of time to progression) were highly correlated. Looking at the predictions of the effect of treatment on log (survival), based on the observed effect of treatment on log (time to progression), the authors concluded that time to progression could be used as a surrogate for survival in advanced ovarian cancer. The effect of treatment could be observed earlier if time to progression were used instead of survival and the effect was also somewhat more pronounced. Hence, a trial that used time to progression would require less follow-up time and fewer patients to establish the statistical significance of a truly superior treatment than a trial that used survival. The gains, however, would be modest because progression was followed by death within 1 year for most patients. Thus in the front-line setting both progression-free survival as a surrogate end point, and overall survival as a true end point are supported by evidence as reasonable primary end points. If progression-free survival is the primary end point, however, and an advantage to a new treatment is shown, information on the survival impact of that treatment will also be an important adjunct to trial results since, regardless of the historical weight of evidence supporting progression-free survival as a primary end point, clinicians will eventually want to know the survival outcome of a particular study. This may be even more important for phase III studies, in which new biological and targeted therapies are investigated, because it is not at all clear whether the relationship between progression-free survival (as a surrogate for overall survival) and overall survival (as the true end point), which is largely based on studies with chemotherapy, also applies for these newer and different forms of therapy. Furthermore, even if further therapy in the control arm at the time of progression dilutes the impact of the new treatment on the end point of overall survival, this would be important to know, because this may suggest that a policy of using the control therapy first, then using another therapy at the time of progression may be as good as using the new treatment in first-line. This all implies that trials should be adequately powered to address both end points with adequate follow-up. If progression-free survival is the primary end point, earlier reporting of data is possible and positive results may lead to earlier adoption of new treatments in some jurisdictions. Nevertheless, this should be followed by the reporting of overall survival data at some stage, to allow a full picture of the policy of using the two treatments to emerge.

End points: second-line phase III studies

For phase III trials in the second-line setting, progression-free survival does not seem to be a good surrogate for survival: there are several examples where progression-free survival was significantly improved, with no survival impact [44–47]. It can be argued that some of these studies were underpowered to detect survival improvements; however, the weight of evidence to consider progression-free survival a surrogate for survival, and thus a primary end point in the second-line setting, is not strong as yet. In the recurrent disease setting, overall survival remains an important primary end point (particularly if more costly or toxic therapy is being offered). Progression-free survival data remain of interest but are unlikely to be sufficiently persuasive to shift practice patterns. Furthermore, since the rationale for treating patients with relapsed disease is a desire to improve symptoms and thus quality of life, an adequate measure of these factors would also be an appropriate primary end point for randomized trials. However, no universally acknowledged and standardized system of symptom measurement and analysis is readily available. GCIG will continue, through its working groups, to build a consensus on how meaningful improvements in disease-related symptoms can be quantified.

End points: maintenance/consolidation phase III studies

A special issue is maintenance and consolidation trials (see also the summary of Workshop C: Integration of new or experimental treatment options and new approaches to clinical trial, this issue). To date, randomized trials with both cytotoxic agents and biological agents are negative, both for progression-free and overall survival, with the exception of the SWOG/GOG trial, which showed a significant difference in progression-free survival in favor of the 12 versus 3 months of maintenance paclitaxel after complete response to platinum and paclitaxel-based chemotherapy [48]. This study was stopped early after a planned interim analysis based on progression-free survival outcomes. Because patients were informed and allowed to continue treatment for 12 months on the 3-month arm, this precluded any meaningful analysis of overall survival benefit. Since trials involving maintenance by definition have longer treatment on the experimental arm as compared with the control, it seems reasonable to expect that progression might be delayed: the real question is whether the prolonged therapy improves survival. Thus, overall survival is the primary end point that should be selected for trials of this design. Interestingly, the next trial in the USA employing prolonged consolidation will randomize patients to no further therapy after front-line chemotherapy versus taxane and will consider overall survival as the primary end point.

End points for interim analysis

The example of the SWOG/GOG trial also raises the issue of early stopping/interim analysis of randomized trials. All such analyses must be pre-specified in the protocol (which was in fact the case in the example cited). However, early stopping for extreme differences (*benefit*) should be based on the primary end point, not an intermediate or surrogate end point since, as was the case for the SWOG/GOG study, to do otherwise may forever impair the ability to perform an analysis of the primary study end point [49]. Therefore, if the primary end point is not overall survival (but, for example, progression-free survival) we suggest that early stopping guidelines for benefit should include both the primary end point and overall survival for the reasons described in the previous section. In cases when the stopping rule is geared to halt the study for reasons of *lack of benefit*, the end point for the analysis may reasonably be either the primary end point or a valid intermediate/surrogate end point.

End points for studies of non-cytotoxic agents

The use of agents that target novel molecular changes in malignancy (as opposed to the usual cytotoxic targets of DNA and tubulin) has raised some interesting questions about study design and end points. Thus, data on non-cytotoxics in ovarian cancer, and in some other tumor types, do not suggest that end points being used in phase I or II trials are any different from those used in trials with cytotoxic agents [21, 50]. While some novel end points, particularly for phase II trials, such as non-progression or imaging measures have been proposed, these are not yet validated and should await this step before application except on an experimental basis.

Once non-cytotoxic drugs are in phase III evaluation, there is no reason to consider end points other than those described above. It will still be important to determine, before changing practice, what the impact of the new agent is on overall, relapse-free or progression-free survival, depending on the

Table 5. Consensus statements in response to question 3

Which are the recommended primary end points for future phase II and randomized phase III clinical trials in ovarian cancer?
The recommended primary end points for future clinical trials in ovarian cancer are:
Phase II screening for activity: response[a] (objective RECIST or GCIG defined CA 125: to be specified in each protocol)
Phase III
Early ovarian cancer: recurrence-free survival (note: recurrence = recurrent disease + death from any cause)
Advanced first-line: both progression-free survival (PFS) and overall survival (OS) are important end points to understand the full impact of any new treatment. Thus either may be designated as the primary end point. Regardless of which is selected, the study should be powered so both PFS and OS can be appropriately evaluated.
Maintenance following first-line: OS[1 minority statement]
Post-recurrence/progression trials: The choice of the primary end point needs to be fully justified with appropriate power calculations. Symptom control/quality of life (for early relapse) and OS (for late relapse) may be the preferred primary end point although PFS should still be used in the assessment of new treatments. Whatever the primary end point, the ability of the study design to detect important differences in survival should be formally addressed.
Interim analysis: end points
Time points for all efficacy analyses should be pre-specified in the protocol
Early stopping/reporting for benefit
Primary end point
If OS is not the primary end point then it is highly recommended that any stopping guidelines include specific criteria for stopping separately for both the primary end point and OS
Early stopping for lack of benefit (in phase III or phase II–III)
Primary or intermediate end points

[a]For non-cytotoxic or biologic agents, other end points such as non-progression, immune response, etc., are being investigated, but are not yet validated.

phase III setting. Thus far, investigators continue to design phase III trials of non-cytotoxic agents using traditional clinical end points [51, 52].

Summary of end point recommendations

With all the above-mentioned considerations working group B formulated the recommendations listed in Table 5.

Acknowledgements

This document summarizes the discussions held in the group B section during the 3rd Ovarian Cancer Consensus Conference, organized by the GCIG and held in Baden-Baden, Germany, 3–5 September 2004. None of the participants in this section indicated any conflict of interest related to the sponsor of this event (Bristol-Myers Squibb).

References

1. Sylvester R. Planning cancer clinical trials. In Buyse ME, Staquet MJ, Sylvester RJ (eds): Oxford Cancer Clinical Trials – Methods and Practice. 47–63. Oxford University Press 1982.
2. Brady MF, Thigpen JT, Vermorken JB, Parmar MKB. Randomised trials in ovarian cancer: trial design considerations. Ann Oncol 1999; 10 (Suppl 1): S75–S82.
3. Cannistra SA. Cancer of the ovary. N Engl J Med 2004; 351: 2519–2529.
4. Berek JS, Bertelsen K, du Bois A et al. Advanced epithelial ovarian cancer: 1998 consensus statements. Ann Oncol 1999; 10 (Suppl 1): S87–S92.
5. Vergote I, De Brabanter J, Fyles A et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. Lancet 2001; 357: 176–182.
6. Zanetta G. The accuracy of staging: an important prognostic determinator in stage I ovarian carcinoma. A multivariate analysis. Ann Oncol 1998; 9: 1097–1101.
7. Eisenhauer EA, Gore M, Neijt JP. Ovarian cancer: should we be managing patients with good and bad prognostic factors in the same manner? Ann Oncol 1999; 10 (Suppl 1): S9–S15.
8. Tropé C, Kaern J, Hogberg T et al. Randomized study on adjuvant chemotherapy in stage I high-risk ovarian cancer with evaluation of DNA-ploidy as prognostic instrument. Ann Oncol 2000; 11: 281–288.
9. Eisenhauer EA, Vermorken JB, van Glabbeke M. Predictors of response to subsequent chemotherapy in platinum pretreated ovarian cancer: an analysis of 704 patients. Ann Oncol 1997; 8: 963–968.
10. van Glabbeke M, Vermorken J, Eisenhauer E. Predictors of survival (OS) and time to progression (TTP) following 2^{nd} and 3^{rd} line treatment in advanced ovarian cancer (OVCA): An analysis of 744 patients (pts) treated with paclitaxel, docetaxel, or high dose epirubicin. Proc Am Soc Clin Oncol 1997; 16: 254a (Abstr 1265).
11. ICON1 and EORTC-ACTION Collaborators. International Collaborative Ovarian Neoplasm trial 1 and Adjuvant Chemotherapy In Ovarian Neoplasm trial: two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma. J Natl Cancer Inst 2003; 95: 105–112.
12. Bell J, Brady M, Lage J et al. A randomized trial of three versus six cycles of carboplatin and paclitaxel as adjuvant treatment in early stage ovarian epithelial carcinoma: a Gynecologic Oncology Group study. Gynecol Oncol 2003; 88: 156 (Abstr 1).
13. International Collaborative Ovarian Neoplasm (ICON1) Collaborators. International Collaborative Ovarian Neoplasm Trial 1: a randomized trial of adjuvant chemotherapy in women with early-stage ovarian cancer. J Natl Cancer Inst 2003; 95: 125–132.
14. Trimbos JB, Vergote I, Bolis G et al. Impact of adjuvant chemotherapy and surgical staging in early-stage ovarian carcinoma: European Organization for Research and treatment of Cancer – Adjuvant Chemotherapy in Ovarian Neoplasm Trial. J Natl Cancer Inst 2003; 95: 113–125.
15. McGuire WP, Hoskins WJ, Brady MF et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and IV ovarian cancer. N Engl J Med 1996; 334: 1–6.
16. Muggia FM, Braly PS, Brady MF et al. Phase III randomized study of cisplatin versus paclitaxel versus cisplatin and paclitaxel in patients with suboptimal stage III or IV ovarian cancer: a Gynecologic Oncology Group study. J Clin Oncol 2000; 18: 106–115.
17. Alberts DS, Liu PY, Hannigan EV et al. Intraperitoneal cisplatin plus intravenous cyclophosphamide versus intravenous cisplatin plus intravenous cyclophosphamide for stage III ovarian cancer. N Engl J Med 1996; 335: 1950–1955.
18. Markman M, Bundy BN, Alberts DS et al. Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the Gynecologic Oncology Group, Southwestern Oncology Group and Eastern Cooperative Oncology Group. J Clin Oncol 2001; 19: 1001–1007.
19. Amstrong DK, Bundy BN, Baergen R et al. Randomized phase III study of intravenous (IV) paclitaxel and cisplatin versus IV paclitaxel, intraperitoneal (IP) cisplatin and IP paclitaxel in optimal stage III epithelial ovarian cancer (OC): a Gynecologic Oncology Group trial (GOG 172). Proc Am Soc Clin Oncol 2002; 21: 201a (Abstr 803).
20. Piccart MJ, Floquet A, Scarfone G et al. Intraperitoneal cisplatin versus no further treatment: 8-year results of EORTC 55875, a randomized phase III study in ovarian cancer patients with a pathologically complete remission after platinum-based intravenous chemotherapy. Int J Gynecol Cancer 2003; 13 (Suppl 2): 196–203.
21. Bookman MA, Darcy KM, Clarke-Pearson D et al. Evaluation of monoclonal humanized anti-HER2 antibody, trastuzumab, in patients with recurrent or refractory ovarian or primary peritoneal carcinoma with overexpression of HER2: a phase II trial of the Gynecologic Oncology Group. J Clin Oncol 2003; 21: 283–290.
22. Royston P, Parmar MKB, Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. Stat Med 2003; 22: 2239–2256.
23. UK Coordinating Committee on Cancer Research (UKCCCR) Ductal Carcinoma In Situ (DCIS) Working Party. Radiotherapy and tamoxifen in women with completely excised ductal carcinoma in situ of the breast in the UK, Australia and New Zealand: randomized controlled trial. Lancet 2003; 362: 95–102.
24. Law MG, Emery S. Selective exclusion of treatment arms in multi-arm randomized clinical trials. Stat Med 2003; 22: 19–30.
25. Hochberg Y, Tamhane AC. Multiple Comparison Procedures. New York: John Wiley and Sons 1987.
26. Dunnett CW, Tamhane AC. A step up multiple test procedure. J Am Stat Assoc 1992; 87: 162–170.
27. Proschan MA, Follman DA. Multiple comparisons with control in a single experiment: Why do we feel differently? Am Stat 1995; 49: 144–149.
28. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: Are we being misled? Ann Int Med 1994; 125: 605–613.

29. Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000; 92: 205–216.
30. Thiesse P, Olivier L, Di Stefano-Louineau D et al. Response rate accuracy in oncology trials: reasons for interobserver variability. J Clin Oncol 1997; 15: 3507–3514.
31. Rustin GJ. Use of CA125 to assess response to new agents in ovarian cancer trials. J Clin Oncol 2003; 21 (Suppl 10): 187–193.
32. Rustin GJS, Quinn M, Thigpen T et al. New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 2004; 96: 487–488.
33. Guastalla J, Vincent P, Le Rol A et al. CA-125 evaluation of chemotherapy response in patients with recurrent ovarian cancer (ROC). Proc Am Soc Clin Oncol 2002; 21: 204a (Abstr 815).
34. Lakusta CM, Atkinson MJ, Robinson JW et al. Quality of life in ovarian cancer patients receiving chemotherapy. Gynecol Oncol 2001; 81: 490–495.
35. du Bois A, Luck HJ, Meier W et al. A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. J Natl Cancer Inst 2003; 95: 1320–1329.
36. Vergote I, Rustin GJS, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst 2000; 92: 1543–1545.
37. Nabholtz JM, Vannetzel JM, Llory JF, Boufette P. Advances in the use of taxanes in the adjuvant therapy of breast cancer. Clin Breast Cancer 2003; 4: 187–192.
38. Piccart MJ, Bertelsen K, James K et al. Randomized trial of cisplatin–paclitaxel versus cisplatin–cyclophosphamide in women with advanced epithelial ovarian cancer: three-year results. J Natl Cancer Inst 2000; 92: 699–708.
39. International Collaborative Ovarian Neoplasm (ICON) Group. Paclitaxel plus carboplatin versus standard chemotherapy with either single-agent carboplatin or cyclophosphamide, doxorubicin, and cisplatin in women with ovarian cancer: the ICON3 randomised trial. Lancet 2002; 360: 505–515.
40. du Bois A, Combe J, Rochon C et al. Epirubicin/paclitaxel/carboplatin (TEC) vs paclitaxel/carboplatin (TC) in first-line treatment of ovarian cancer (OC) FIGO stage IIB–IV. An AGO-GINECO Intergroup phase III trial. J Clin Oncol 2004; 22: 450 (Abstr 5007).
41. De Placido S, Scambia G, Di Vagno G et al. Topotecan compared with no therapy after response to surgery and carboplatin/paclitaxel in patients with ovarian cancer: Multicenter Italian Trials in Ovarian cancer (MITO-1) randomized study. J Clin Oncol 2004; 22: 2635–2642.
42. Buyse M, Molenberghs G, Burzykowski T et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1: 49–67.
43. Ovarian Cancer Meta-analysis Project. Cyclophosphamide plus cisplatin versus cyclophosphamide, doxorubicin, and cisplatin chemotherapy of ovarian carcinoma: a meta-analysis. J Clin Oncol 1991; 9: 1668–1674.
44. Eisenhauer EA, ten Bokkel Huinink WW, Swenerton KD et al. European–Canadian randomized trial of paclitaxel in relapsed ovarian cancer: high-dose versus low-dose and long versus short infusion. J Clin Oncol 1994; 12: 2654–2666.
45. Pfisterer J, Plante M, Vergote I et al. Gemcitabine/carboplatin (GC) vs carboplatin (C) in platinum sensitive recurrent ovarian cancer (OVCA). Results of a Gynecologic Cancer Intergroup randomized phase III trial of the AGO OVAR, the NCIC CTG and the EORTC GCG. J Clin Oncol 2004; 22: 450 (Abstr 5005).
46. du Bois A, Meier W, Luck HJ et al. Chemotherapy versus hormonal treatment in platinum- and paclitaxel-refractory ovarian cancer: a randomized trial of the German Arbeitsgemeinschaft Gynaekologische Onkologie (AGO) Study Group Ovarian Cancer. Ann Oncol 2002; 13: 251–257.
47. ten Bokkel Huinink W, Gore M, Carmichael J et al. Topotecan versus paclitaxel for the treatment of recurrent epithelial ovarian cancer. J Clin Oncol 1997; 15: 2183–2193.
48. Markman M, Liu PY, Wilczynski S et al. Phase III randomized trial of 12 versus 3 months of maintenance paclitaxel in patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest Oncology Group and Gynecologic Oncology Group trial. J Clin Oncol 2003; 21: 2460–2465.
49. Cannistra S. The eithics of early stopping rules: who is protecting whom? J Clin Oncol 2004; 22: 1542–1545.
50. Perulekar WR, Eisenhauer EA. Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: theory and practice. J Natl Cancer Inst 2004; 96: 990–997.
51. Hall GD, Brown JM, Coleman RE et al. Maintenance treatment with interferon for advanced ovarian cancer: results of the Northern and Yorkshire gynaecology group randomised phase III study. Br J Cancer 2004; 91: 621–626.
52. Biagi JJ, Eisenhauer EA. Systemic treatment policies in ovarian cancer: the next 10 years. Int J Gynecol Cancer 2003; 13 (Suppl 2): 231–240.